

Large attractors in cooperative bi-quadratic Boolean networks. Part I.

German A. Enciso* and Winfried Just†

February 2, 2008

Abstract

Boolean networks have been the object of much attention, especially since S. Kauffman proposed them in the 1960's as models for gene regulatory networks. These systems are characterized by being defined on a Boolean state space and by simultaneous updating at discrete time steps. Of particular importance for biological applications are networks in which the indegree for each variable is bounded by a fixed constant, as was stressed by Kauffman in his original papers.

An important question is which conditions on the network topology can rule out exponentially long periodic orbits in the system. In this paper, we consider systems with positive feedback interconnections among all variables (known as cooperative systems), which in a continuous setting guarantees a very stable dynamics. We show that for an arbitrary constant $0 < c < 2$ and sufficiently large n there exist n -dimensional cooperative Boolean networks in which both the indegree and outdegree of each variable is bounded by two, and which nevertheless contain periodic orbits of length at least c^n . In Part II of this paper we will prove an inverse result showing that any system with such a dynamic behavior must in a sense be similar to the example described.

Keywords: Boolean networks, monotone systems, periodic solutions, mathematical biology, gene regulatory networks

AMS Subject Classification: 34C12, 39A11, 92B99.

The concept of a *Boolean network* was originally proposed in the late 1960's by Stuart Kauffman to model gene regulatory behavior at the cell level [12, 13]. This type of modeling can sometimes capture the general dynamics of continuous systems in a simplified framework, e.g. without the choice of specific nonlinearities or parameter values; see for instance [1]. Boolean networks are known and used in several other disciplines such as

*Mathematical Biosciences Institute, Ohio State University, and Harvard Medical School, Department of Systems Biology.

†Department of Mathematics, Ohio University.

This material is based upon work supported by the National Science Foundation under Agreement No. 0112050 and by The Ohio State University.

electrical engineering, computer science, and control theory, and analogous definitions are known under various names such as sequential dynamical systems [7] or Boolean difference equations [6].

An important class of *continuous* dynamical systems is that of so-called *monotone systems*, which can be roughly characterized by the absence of negative feedback interactions [2, 15]. A special case is that of *cooperative systems*, in which there are no direct inhibitory interactions between any two variables. Monotone and cooperative systems have been used as a modeling tool for gene regulatory systems, for instance in [3]. The assumption of monotonicity is a stringent condition which ensures that the system behavior is remarkably stable: for instance, under mild additional assumptions the generic solution of a monotone dynamical system must converge towards an equilibrium.

In the Boolean case, the class of cooperative systems can be described as that corresponding to maps that can be expressed using only AND and OR gates, i.e. with no use of negations. This can be easily seen by considering the disjunctive normal form of the Boolean maps.

An important question in the study of cooperative Boolean networks is whether some of the stability properties of continuous cooperative systems have analogues in the Boolean case. For instance, does the assumption of cooperativity by itself limit the length of the longest cycle in an n -dimensional Boolean system? It was shown recently through simulations that random Boolean systems tend to have shorter periodic cycles if they are cooperative, or even if they are close to cooperative in the sense of having few negative feedback interactions; see [16], and also [9, 19]. Nevertheless, a straightforward use of Sperner's theorem shows that a cooperative n dimensional Boolean system can have a cycle of length close to 2^n for large n , see [8] and more recently [11, 17].

One would like to know which additional assumptions rule out exponentially long periodic orbits in cooperative Boolean systems. In [11] suitable adaptations of the notion of *strong cooperativity* [15] to Boolean systems were found that limit the length of periodic orbits to $2^{\sqrt{n} \log n (1+o(1))}$ or even to n , the dimension of the system. In the present manuscript we follow up on this question by considering a different class of cooperative Boolean systems in which both the indegree and the outdegree of the associated digraph is bounded.

We need some definitions. An n -dimensional *Boolean dynamical system* or *Boolean network* is a pair (Π, g) , where $\Pi = \{0, 1\}^n$ and $g : \Pi \rightarrow \Pi$. A state $s(t)$ at time t will be denoted by $s(t) = [s_1(t), \dots, s_n(t)]$, or simply $s = [s_1, \dots, s_n]$ if time-dependency is ignored. We will have

$$s(t+1) = g(s(t)). \quad (1)$$

The *cooperative order* on Π is the partial order relation defined by $s \leq r$ iff $s_i \leq r_i$ for all $i \in \{1, \dots, n\}$. The system is *cooperative* if $s(t) \leq r(t)$ implies $s(t+1) \leq r(t+1)$.

We associate a directed graph D with vertex set $\{1, \dots, n\}$ with the system. A pair $\langle i, j \rangle$ is in the arc set of D iff there exist states $s, r \in \Pi$ such that $s_i < r_i$ and $s_k = r_k$ for all $k \neq i$ with the property that $(g(s_i))_j < (g(r_i))_j$. We will say that the system is *bi-quadratic* if both the indegree and the outdegree of all vertices in D is at most two.

Already in his 1969 papers [12, 13], Kauffman focused his attention on Boolean networks where every variable can only be directly affected by a fixed number K of other variables. In the digraph associated to the network, this corresponds to limiting the indegree of every node to (at most) K . This corresponds to empirical findings about actual gene regulatory networks which show that most genes are directly regulated by a small number of proteins in a scale-free manner [4, 18]. Other studies of biochemical networks show that only very few nodes are involved in the regulation of other chemicals. Thus large subnetworks of most biochemical networks of interest will also have the property that the outdegree of each node is bounded by a small integer. Bi-quadratic Boolean networks satisfy both of these restrictions with $K = 2$. Random Boolean networks with $K = 2$ have been extensively studied and tend to have dynamics in the *ordered regime*, which is characterized, among other properties, by the absence of exponentially long attractors (see [14] for a review). Thus it becomes a natural question whether one can prove, for cooperative bi-quadratic Boolean networks, a subexponential bound on the length of their periodic orbits, or at least a bound of the form c^n for some constant $c < 2$. The following theorem shows that this is not the case.

Theorem 1 *Let $c < 2$ be arbitrary. Then for some sufficiently large n there exists an n -dimensional, bi-quadratic cooperative Boolean network which contains a periodic orbit of length at least c^n . Moreover the digraph D associated with this network is strongly connected.*

The last sentence of Theorem 1 is of interest in connection with the results in [11]. There, we define a local version D_s of D for every state s as follows: A pair $\langle i, j \rangle$ is in the arc set of D_s iff there exist a state $r \in \Pi$ such that either $s_i < r_i$ while $s_k = r_k$ for all $k \neq i$, and we have $(g(s_i))_j < (g(r_i))_j$, or $r_i < s_i$ while $s_k = r_k$ for all $k \neq i$, and we have $(g(r_i))_j < (g(s_i))_j$. It is shown that if X is a periodic orbit of an n -dimensional cooperative Boolean system such that D_s is strongly connected for every $s \in X$, then $|X| \leq n$ (Theorem 25 of [11]).

The proof of Theorem 1 uses a construction similar to a small Turing machine operating on a long circular tape. In part II of this paper we will show that if c is sufficiently close to 2, then all n -dimensional bi-quadratic cooperative Boolean networks with periodic orbits of length $\geq c^n$ must contain a relatively small subsystem that can be considered a Turing machine operating on one or more tapes that retain the values of all other variables.

The remainder of this note is organized as follows: In Section 1 we introduce the main idea of the construction, but without requiring the system to be cooperative and bi-quadratic. In Section 2 we show how to modify the construction so that the network will also be cooperative, bi-quadratic and will have a strongly connected digraph.

1 A Simple Counting Model

In this subsection we consider a conceptual model of a (not necessarily bi-quadratic or cooperative) Boolean network with periodic orbits of length 2^N , for arbitrary $N > 0$.

We also discuss the problems that are involved in constructing such a network under the restrictions of Theorem 1. Consider the states s_1, \dots, s_N , and the system defined by

$$\begin{aligned} s_i(t) &:= s_{i+1}(t-1), \quad i = 1, \dots, N-1, \\ s_N(t) &:= \gamma(s_1(t-1), \text{mode}(t-1)). \end{aligned} \quad (2)$$

One can think of γ on a conceptual level as a Turing machine operating on variables numbered $i = 1, \dots, N$ whose values are written on a circular tape. The variable *mode* can have one of two possible values for every t , namely *mode* = *rotate*, and *mode* = *switch*, and the function γ is defined by

$$\begin{aligned} \gamma(x, \text{rotate}) &= x, \\ \gamma(x, \text{switch}) &= 1 - x. \end{aligned} \quad (3)$$

Thus while *mode*(t) = *rotate*, iterating this machine will cyclically rotate the values of s_1, \dots, s_N . Whenever *mode* = *switch*, the machine also will rotate the variable values, but it will invert them at the site s_N .

Now let us define the value of the variable *mode*, in such a way that this machine behaves like a counter in base two. Let us require that at the times $t = 0, N, 2N, 3N, \dots$, *mode*(t) = *switch*. For all other times t , define

$$\text{mode}(t) := \begin{cases} \text{mode}(t-1), & \text{if } s_1(t-1) = 1, \\ \text{rotate}, & \text{if } s_1(t-1) = 0. \end{cases} \quad (4)$$

Thus the model turns into *switch* mode exactly at the times $t = 0, N, 2N, \dots$, and it only returns back to *rotate* mode after $s_1(t_1) = 0$ for some $t_1 > t$. The following lemma shows in what way this machine is a counter: if the states of the system encode numbers in binary format appropriately, then N iterations are equivalent to the addition of one unit modulo 2^N .

Lemma 2 *Given any state s of the model, define $\alpha(s) := s_1 2^0 + s_2 2^1 + \dots + s_N 2^{N-1}$. Then $\alpha(s(N)) = \alpha(s(0)) + 1 \bmod 2^N$.*

Proof: Consider an initial state $s(0)$ and let $j \geq 0$ be such that $s_i(\eta) = 1$, for $1 \leq \eta \leq j < N$, and $s_{j+1}(0) = 0$. Note that $\alpha(s(0)) < 2^N - 1$ in this case. We have *mode*(0) = *switch* by the definition above (4). By (2), $s_1(\eta) = 1$ for $0 \leq \eta \leq j-1$, $s_1(j) = 0$. Therefore *mode*(η) = *switch*, for $1 \leq \eta \leq j$, and *mode*($j+1$) = \dots = *mode*($N-1$) = *rotate*. At time $t = N$, the variable values have completed a full rotation and returned to their starting points, except that $s_\eta = 0$ for $1 \leq \eta \leq j$, $s_{j+1} = 1$, and s_{j+2}, \dots, s_N are unchanged. Clearly $\alpha(s(N)) = \alpha(s(0)) + 1$ in this case.

It remains to show the result for the case $j = N$, i.e. $s_i(0) = 1$, for every $i = 1, \dots, N$. In that case *mode*(0) = *mode*(1) = \dots = *mode*($N-1$) = *switch* by (2) and (4). In this way every value of the system is inverted at s_1 from 1 to 0, so that $s_i(N) = 0$ for $i = 1 \dots N$. Therefore $\alpha(s(N)) = 0 = \alpha(0) + 1 \bmod 2^N$. \square

Corollary 3 *The network given by equations (2), (3), (4), contains a periodic cycle of length at least 2^N .*

Proof: Since the variable *mode* is reset to *switch* for $t = 0, N, 2N, \dots$, Lemma 2 applies at each of these time points. Therefore one can start with $s(0) = 0$, and apply Lemma 2 successively to reach states $s(0), s(N), s(2N), \dots, s((2^N - 1)N)$, which are all different from each other. \square

Importantly, the function γ negates the values of the input x in switching mode. This appears to be an essential non-monotonic component (or negative feedback) of this system. Nevertheless, it is shown below that in fact one can rewrite our system in such a way that the resulting system is cooperative.

1.1 A Generalized Counter

Before proceeding with the proof of the main result, consider the following generalization of the simple counter above. Instead of individual Boolean values, each variable s_i is now considered to be a vector with $l > 1$ Boolean entries, $s_i = (s_i^1, \dots, s_i^l)$. We will treat s_i as a binary code for a nonnegative integer $< 2^l$. At each time t , the system continues to be in one of two modes $mode(t) = \text{switch}$ or $mode(t) = \text{rotate}$, but the function γ is now replaced with a vector function Γ which we describe in the next paragraph.

As before, when $mode = \text{rotate}$ we let $\Gamma(x, mode) := x$. When $mode = \text{switch}$, and given $x = (x^l, x^{l-1}, \dots, x^1) \neq (1, \dots, 1)$, let j be such that $x^\eta = 1$ for $1 \leq \eta \leq j < l$, $x^{j+1} = 0$. Define y by letting $y^\eta := 0$ for $1 \leq \eta \leq j$, letting $y^{j+1} := 1$, and $y^\eta := x^\eta$ for $j+1 < \eta \leq l$. Set $\Gamma(x, \text{switch}) := y$. If $x = (1, \dots, 1)$, set $\Gamma(x, \text{switch}) := (0, \dots, 0)$. In other words, the function $\Gamma(x, \text{switch})$ is defined as the addition of 1 to the vector x , in base 2 and modulo 2^l .

We define the generalized system

$$\begin{aligned} s_i(t) &:= s_{i+1}(t-1), \quad i = 1, \dots, N-1, \\ s_N(t) &:= \Gamma(s_1(t-1), mode(t-1)), \end{aligned} \tag{5}$$

where Γ is defined as above. The variable $mode(t)$ has the value *switch* for $t = 0, N, 2N, \dots$ and for other values of t :

$$mode(t) := \begin{cases} \text{switch}, & \text{if } s_1(t-1) = (1, \dots, 1), \\ \text{rotate}, & \text{otherwise.} \end{cases} \tag{6}$$

Lemma 4 *The network defined by equations (5), (6) contains a periodic cycle of length at least 2^{Nl} .*

Proof: For $(x^l, \dots, x^1) \in \{0, 1\}^l$, define $\beta(x) := x^1 2^0 + x^2 2^1 + \dots + x^l 2^{l-1}$. Note that $\beta(\Gamma(x, \text{switch})) = \beta(x) + 1 \pmod{2^l}$. We follow an argument very analogous to Lemma 2 and Corollary 3. Let $\alpha(s) := \beta(s_1)(2^l)^0 + \beta(s_2)(2^l)^1 + \dots + \beta(s_N)(2^l)^{N-1}$. Thus the vector $(\beta(s_1), \dots, \beta(s_N))$ can be regarded as the representation of $\alpha(s)$ in base 2^l .

As in the proof of Lemma 2, consider an initial state $s(0)$, and let $j \geq 0$ be such that $s_\eta(0) = (1, \dots, 1)$, for $1 \leq \eta \leq j < N$, and $s_{j+1}(0) \neq (1, \dots, 1)$. As before, we have $\text{mode}(\eta) = \text{switch}$ for $0 \leq \eta \leq j$, and $\text{mode}(j+1) = \dots = \text{mode}(N-1) = \text{rotate}$. At time $t = N$ we have $s_\eta = (0, \dots, 0)$ for $1 \leq \eta \leq j$, as well as $\beta(s_{j+1}) = \beta(s_{j+1}(0)) + 1$, and s_{j+2}, \dots, s_N are unchanged from $t = 0$. Clearly $\alpha(s(N)) = \alpha(s(0)) + 1$.

In the case that $s_i(0) = (1, \dots, 1)$ for every $i = 1, \dots, N$, it follows as before that $\text{mode}(0) = \text{mode}(1) = \dots = \text{mode}(N-1) = \text{switch}$. Therefore $s_i(N) = (0, \dots, 0)$ for $i = 1 \dots N$, and $\alpha(s(N)) = 0$.

Repeating this process for $s(0) \equiv 0$ and $t = N, 2N, \dots$, as in Corollary 3, one finds states s of the system such that $\alpha(s) = 1, 2, \dots$, and which are therefore pairwise different. When $s_i = (1, \dots, 1)$ for all i , that is, when $\alpha(s(t)) = (2^t)^N - 1$, this process reverts to $\alpha(s(t+N)) = 0$. \square

2 A Cooperative Counter

In this section we carry out a construction which is analogous to that in Section 1, but in which the underlying Boolean network is cooperative, bi-quadratic, and has a strongly connected digraph. We will need to define some auxiliary Boolean networks with designated input and output variables.

Throughout this section let $L > 0$ be an arbitrary even number, and consider the set $A := \{(r_1, \dots, r_L) \in \{0, 1\}^L \mid s_1 + \dots + s_L = L/2\}$. Define the special sequences $\text{START} = (1, \dots, 1, 0, \dots, 0)$, i.e. $L/2$ ones followed by $L/2$ zeros, and similarly $\text{ACTIVE} = (0, \dots, 0, 1, \dots, 1)$.

Lemma 5 *Let $g : A \rightarrow A$ be an arbitrary function. There exists a Boolean network B with input vectors $a = (a_1, \dots, a_L)$, $d = (d_1, d_2)$, and output vector $c = (c_1, \dots, c_L)$, such that for some fixed $m > 0$ the following equation holds for every t and $a(t) \in A$, regardless of the initial condition of B :*

$$c(t+m) := \begin{cases} a(t), & \text{if } d(t) = (0, 1), \\ g(a(t)), & \text{if } d(t) = (1, 0). \end{cases} \quad (7)$$

Furthermore, the network B is cooperative, every node of its associated digraph has in- and outdegree of at most 2, and the indegree (outdegree) of every designated input (output) variable is zero.

Proof: Define the set $\hat{A} := A \times \{(0, 1), (1, 0)\}$, and the function $G : \hat{A} \rightarrow A$ by $G(x, (1, 0)) := g(x)$, $G(x, (0, 1)) := x$, for arbitrary $x \in A$. Since \hat{A} is an unordered set, G can be extended to a cooperative function $G : \{0, 1\}^{L+2} \rightarrow \{0, 1\}^L$; see [11]. The result will follow from building a Boolean network that computes the function G .

Consider a fixed component $G_i : \{0, 1\}^{L+2} \rightarrow \{0, 1\}$ of G . By the cooperativity of this function, one can write it in the normal form $G_i(y_1, \dots, y_{L+2}) = \Psi_1^i(y_1, \dots, y_{L+2}) \vee$

$\dots \vee \Psi_{k_i}^i(y_1, \dots, y_{L+2})$, where each Ψ_j^i is the conjunction of a number of variables, i.e. $\Psi_j^i(y_1, \dots, y_{L+2}) = y_{\alpha_{1i}} \wedge \dots \wedge y_{\alpha_{ji}}$. This suggests a way of computing G_i : define Boolean variables $\psi_j^i(t) := \Psi_j^i(y(t-1))$, and then let $G_i(t) := \psi_1^i(t-1) \vee \dots \vee \psi_{k_i}^i(t-1)$. Repeating this procedure for all components of G yields a Boolean network which computes G in $m = 2$ steps, and which is cooperative and has indegree (outdegree) zero for every input (output).

In order to satisfy the condition that every node have in- and outdegree of at most two, we need to modify this construction by introducing additional variables. First, note that the outdegree of every input y_i can be very large. One can define two additional variables which simply copy the value of $y_i(t)$, then four variables that copy the value of the previous two, etc. This procedure is repeated for each y_i so that at least as many copies of each variable are present as appear in the expressions of all ψ_j^i . A similar cascade can be used to define each ψ_j^i and G_i so that each indegree is at most two. If $\psi_i^j = y_{\alpha_1} \wedge y_{\alpha_2} \wedge y_{\alpha_3}$, say, then one can define $z_1(t) := y_{\alpha_1}(t-1)$, $z_2(t) := y_{\alpha_2}(t-1) \wedge y_{\alpha_3}(t-1)$, $\psi_i^j(t) := z_1(t-1) \wedge z_2(t-1)$. Similarly for longer disjunctions and each ψ_j^i and also similarly for G_i , in which case \wedge is replaced by \vee at each step. This produces a computation of G_i in m_i steps for each i . Finally, after introducing further additional variables at each component i if necessary to compensate for unequal lengths of the expressions for ψ_j^i , the Boolean vector $G(y_1, \dots, y_{L+2})$ can be computed in exactly $m = \max(m_1, \dots, m_L)$ steps. \square

Remark: Without loss of generality, we can assume that for every state variable s in the network B , there exists some input variable d_i or a_i and a directed path from this input towards s . This is because if that wasn't the case, one could delete s from the system without altering equation (7). Cooperativity of G implies that $G(0, \dots, 0) = (0, \dots, 0)$ and $G(1, \dots, 1) = (1, \dots, 1)$ [11]; therefore each G_i is non-constant and no output variable will be deleted. Similarly, it will be assumed that for every state variable s , there exists an output variable c_i such that there is a directed path from s to c_i .

Lemma 5 can be used to compute a function g which will be used in a way analogous to γ in equation (2). Similarly, we need to construct a 'switch' to determine when to turn the system into *rotate* mode, which is provided by Lemma 6 below. Note that Lemma 5 cannot be used for this purpose because the desired output depends not only on the current state of the input $p(s)$ but on the whole history (of unknown length) of the input sequence since the last time when $p(s)$ took the value *START*.

Lemma 6 *There exists $\mu > 0$ and a Boolean network D with input vector $p = (p_1, \dots, p_L)$, and output vector $q = (q_1, q_2)$, such that the following holds for any initial condition of D . Consider any sequence of inputs $p(0), p(1), \dots, p(M)$, $M > 1$, such that*

- i) $p(s) \in A$, for $0 \leq s \leq M$,*
- ii) $p(0) = \text{START}$, and*
- iii) $p(s) \neq \text{START}$, for $0 < s \leq M$.*

Let $j \geq 0$ be such that $p(s) = \text{ACTIVE}$ for $1 \leq s \leq j$, $p(j+1) \neq \text{ACTIVE}$ (or $p(1) = \dots = p(M) = \text{ACTIVE}$ and $j = M$). Then

$$q(s) = \begin{cases} (1, 0), & \mu \leq s \leq \mu + j, \\ (0, 1), & \mu + j < s \leq \mu + M, \end{cases} \quad (8)$$

Furthermore, the network B is cooperative, every node of its associated digraph has in- and outdegree of at most 2, and the indegree (outdegree) of every designated input (output) variable is zero.

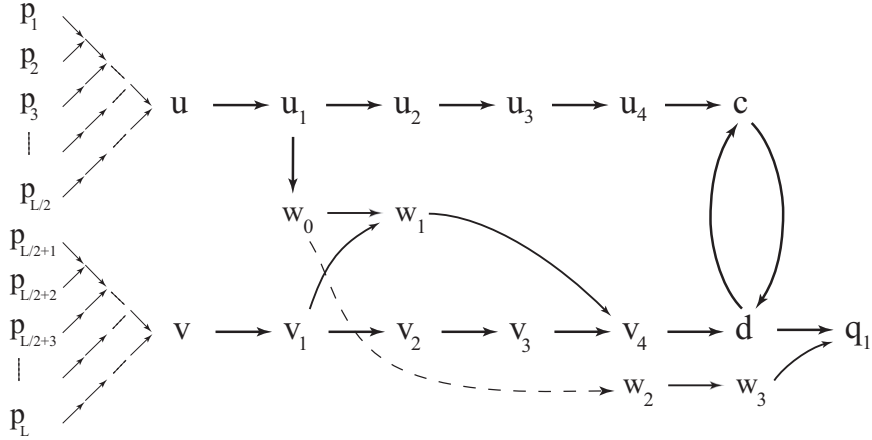


Figure 1: The digraph of the network D which is used to compute the output q_1 from the input p . The formulas for each interaction (i.e. \wedge, \vee) as well as the dependencies of u_2 on u and v_2 on v are omitted in this figure.

Proof: The idea for this proof is the simple system $c(t) = u(t-1) \vee d(t-1)$, $d(t) = v(t-1) \wedge c(t-1)$, with inputs u, v . This switch is turned on by letting both inputs $u = 1$ and $v = 1$ for a short time, after which u can be turned to 0 while v is left equal to 1. After letting $v = 0$ for a short time, the switch resets and doesn't restart even if $v = 1$ again.

Let $t = 0$ without loss of generality, the more general case being completely analogous. For the sake of clarity assume for now that $0 < j < M$, but the same construction allows for $j = 0$ and $j = M$ as described below. See Figure 1 which displays the circuit described below. Define for the moment $u(t) := p_1(t-1) \wedge \dots \wedge p_{L/2}(t-1)$, $v(t) := p_{L/2+1}(t-1) \wedge \dots \wedge p_L(t-1)$ (a modification of this definition with additional variables and indegree two is displayed in the figure and described below). Thus $u(s) = 1$ if and only if $p(s-1) = \text{START}$, and $v(s) = 1$ if and only if $p(s-1) = \text{ACTIVE}$, since by assumption $p(s) \in A$.

Define

$$\begin{aligned} u_1(t) &:= u(t-1), \quad u_2(t) := u(t-1) \vee u_1(t-1), \quad u_3(t) := u_2(t-1), \quad u_4(t) := u_3(t-1), \\ v_1(t) &:= v(t-1), \quad v_2(t) := v(t-1) \wedge v_1(t-1), \quad v_3(t) := v_2(t-1), \quad v_4(t) := v_3(t-1) \vee w_1(t-1), \end{aligned}$$

$$\begin{aligned}
w_0(t) &:= u_1(t-1), \quad w_1(t) := w_0(t-1) \wedge v_1(t-1), \\
c(t) &:= u_4(t-1) \vee d(t-1), \quad d(t) := v_4(t-1) \wedge c(t-1).
\end{aligned}$$

(Intuitively, u_4 is a time-transposed copy of u where every 1 has been doubled due to the feed-forward loop at u_2 . Also, v_4 is similar to a time-transposed copy of v where every 0 has been doubled - the auxiliary variables w_i only play a role at a single time step as described below. The loop $c \leftrightarrow d$ forms the core of the switch in the system.)

A simple calculation shows that $u_4(4) = u_4(5) = 1$, $u_4(s) = 0$ for $5 < s \leq M+4$. On the other hand, since $v(1) = 0, v(2) = \dots = v(1+j) = 1, v(2+j) = 0$, we infer that $v_2(2) = v_2(3) = 0$, $v_2(s) = 1$ for $3 < s \leq 2+j$, $v_2(3+j) = v_2(4+j) = 0$. It follows that $w_1(3) = 0$ (since $v_1(2) = 0$), and that $w_1(4) = 1$ if and only if $v_1(3) = 1$ (since $w_0(3) = 1$). This in turn holds since $j > 0$. Also, $w_1(s) = 0$ for $s > 4$.

We use the data for w_1 and v_3 to compute the values of v_4 . From $w_1(3) = v_3(3) = 0$, it follows that $v_4(4) = 0$. From $w_1(4) = 1$ it follows that $v_4(5) = 1$, and using v_3 we similarly infer that $v_4(s) = 1$ for $s = 4 < s \leq 4+j$. Also, $v_4(5+j) = v_4(6+j) = 0$.

We conclude that $c(5) = 1, d(5) = 0$, regardless of the values of c, d at earlier time steps. Since $j > 0$, one has $c(6) = 1, d(6) = 1$, and in general $c(s) = d(s) = 1$ for $5 < s \leq 5+j$. Then $c(6+j) = 1, d(6+j) = 0, c(s) = d(s) = 0$, for $7+j \leq s \leq 5+M$, and $d(6+M) = 0$.

In particular $d(s) = 1$ for exactly j time steps, $5 < s \leq 5+j$, and then $d(s) = 0$ for $6+j \leq s \leq 6+M$. Since we want the variable q_1 to be equal to 1 during exactly $j+1$ time steps, we define the additional variables

$$w_2(t) := w_0(t-1), \quad w_3(t) := w_2(t-1), \quad q_1(t) := w_3(t-1) \vee d(t-1).$$

Calculating that $w_3(5) = 1, w_3(s) = 0$ for $5 < s \leq 5+M$, we have $q_1(s) = 1, 6 \leq s \leq 6+j$, and $q_1(s) = 0, 6+j < s \leq 7+M$.

In order to define the variable q_2 , it suffices to make a construction dual to the previous one (recall that simply negating q_1 is not permitted). That is, define $\hat{u}(t) := p_{L/2+1}(t-1) \vee \dots \vee p_L(t-1)$, and $\hat{v}(t) := p_1(t-1) \vee \dots \vee p_{L/2}(t-1)$, in such a way that $\hat{u}(s) = 0$ if and only if $p(s-1) = \text{START}$, and $\hat{v}(s) = 0$ if and only if $p(s-1) = \text{ACTIVE}$. Define variables \hat{u}_1, \hat{v}_1 etc. similarly as above, except that every \wedge in the function definition is replaced by \vee and vice versa. Then it will necessarily follow that $q_2 = \neg q_1$ on the interval $6 \leq s \leq 6+M$. Using the value $\mu = 6$, equation (8) is satisfied.

The case $j = 0$ is very similar as above. In this case $w_1(4) = 0$ (instead of 0 for $j > 0$), $v_4(4) = v_4(5) = 0$, and therefore $d(s) = 0$ on all $6 \leq s \leq M+6$. Thus $q_1(6) = 1$, and $q_1 = 0$ for larger values of s . In the case $j = M$, one can compute $v_4(s) = 1$ for $5 \leq s < M+5$. This allows the variables $c(s), d(s)$ to remain equal to 1 up to and including $s = M+5$. Therefore $q_1(s) = 1$ up to and including $s = 6+M$.

Notice that this system is cooperative, and that all in- and outdegree requirements are satisfied except for the indegree of the variables u, v, \hat{u}, \hat{v} . These terms can now be replaced in a routine manner by a cascade of variables (see Figure 1), in such a way that $u(s) = 1$ if and only if $p(s-\tau) = \text{START}$, etc., for some $\tau > 1$. This will increase the delay μ but leave the computations and the other properties of this system unchanged. \square

We are ready for the construction of the cooperative counter described in the introduction. This Boolean network is designed to replicate the behavior of the system described by equations (5), (6), while ensuring its cooperativity. In order to do so, we let $l > 0$ be arbitrary and $L > 0$ be an even positive integer, which is large enough that there exists an injective function $\chi : \{0, 1\}^l \rightarrow A$, where A is defined as above. The cooperative network will contain L -dimensional vectors $r_i = (r_i^1, \dots, r_i^L)$, with values in A , which will be considered as proxy for states $s_i = \chi^{-1}(r_i)$ of the system (5), (6).

We require that $\chi(1, \dots, 1) = \text{ACTIVE}$, and that $\text{START} \notin \text{Im}(\chi)$ (see the definitions of START and ACTIVE above). This is possible if L is large enough so that $\binom{L}{L/2} > 2^l$. We also let $\chi(0, \dots, 0) = (1, 0, \dots, 1, 0)$, and $\chi(0, 0, \dots, 0, 1) = (0, 1, 0, 1, \dots, 0, 1)$. Having defined χ , we define $g : A \rightarrow A$ as $g(r) := \chi(\Gamma(\chi^{-1}(r)))$, for $r \in \text{Im}(\chi)$, $g(r) = r$ for all other $r \in A$. The function Γ is defined as in Section 1. In particular, $g(\text{START}) = \text{START}$.

Using the function g defined above, we consider the cooperative networks B and D from Lemmas 5 and 6. Recall that B (D) has variables a, d (p) which are specifically designated as inputs, a variable c (q) specifically designated as output, and a ‘processing delay’ m (μ). The cooperative network, which will be denoted by S , is defined by B and D , together with the equations

$$\begin{aligned} r_i(t) &:= r_{i+1}(t-1), \quad i = m+2, m+3, \dots, N, \\ r_{N+1}(t) &:= c(t-1), \end{aligned} \tag{9}$$

and

$$\begin{aligned} a(t) &:= r_{m+2}(t-1), \\ d(t) &:= q(t-1), \\ p(t) &:= r_{m+\mu+2}(t). \end{aligned} \tag{10}$$

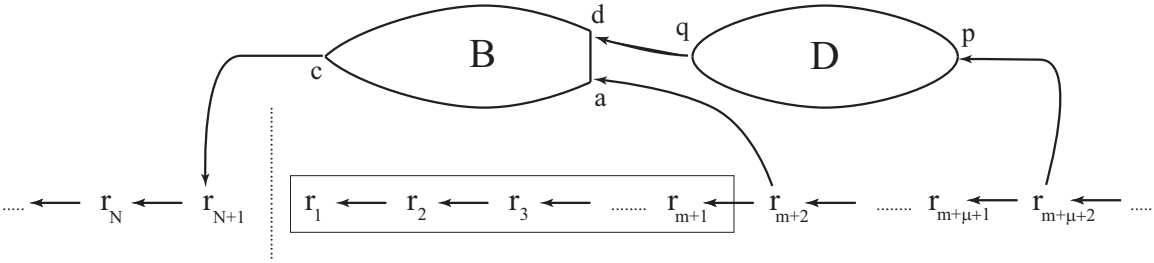


Figure 2: The network interconnections of the system S given by B , D , and equations (9), (10). The variables r_1, \dots, r_{m+1} are displayed in a box to indicate that they are not part of S but only included in the proof of Theorem 8.

See Figure 2 for an illustration. Since both of the subnetworks used in the construction of this system contain only the Boolean operators \wedge, \vee in their expression (and no negations),

it follows from (9) and (10) that the same is the case for the full network, hence the system is cooperative.

Proposition 7 *The digraph of the Boolean network S is strongly connected and bi-quadratic.*

Proof: The fact that every in- and outdegree is at most 2 follows directly from equations (9), (10) and Lemmas 5 and 6, taking into account that the indegree (outdegree) of every input (output) variable is zero within their respective subnetwork. See also Figure 2.

In order to show the strong connectivity of the digraph, first we show that there exists a directed path from every node in the network to the node q_1 , the first component in the output of D . It is clear from the circuit defining D that every input variable p_i has a path connecting to q_1 (the first $L/2$ components of p through the variables u, u_1, \dots and the last $L/2$ components through v, v_1, \dots). Therefore every component of every variable r_i can reach q_1 as well. By the remark after Lemma 5, the same applies to every variable of c , and thus to every variable in the subnetwork B . Thus the same applies also to q_2 , and hence to every state in the subnetwork D .

Now we show that there exists a path from q_1 to every node in the network. Suppose first that there exists c_j such that neither d_1 or d_2 contains a path towards c_j . This would imply that $g_j(x) = x_j$ for every argument $x \in A$, by equation (7). But we have

$$g(1, 0, \dots, 1, 0) = \chi(\Gamma(0, \dots, 0)) = \chi(0, \dots, 0, 1) = (0, 1, 0, 1, \dots, 0, 1),$$

which is a contradiction. Thus for every j , there exists a path from either d_1 or d_2 to c_j (and therefore from q_1 or q_2 to c_j).

Since there exists a path from q_1 to q_2 , it follows that there is a path from q_1 to every c_j . Thus every component of every state r_i , p , and a can be reached by a path from q_1 . Every state in B can be reached from d_1 and hence q_1 , once again by the remark after Lemma 5; the same applies to q_2 , and every state in the subnetwork D . \square

Theorem 8 *Let $L > 0$ be an even number such that $\binom{L}{L/2} > 2^l$. Then system S has a periodic orbit of length greater than or equal to 2^l .*

Proof: For the purposes of this proof, we extend the system with the auxiliary variables r_1, \dots, r_{m+1} , defined by $r_i(t) := r_{i+1}(t-1)$, $i = 1, \dots, m+1$; see Figure 2. These variables cannot change the length of the original system's periodic orbits (since they don't feed back into it), but they can nevertheless be used for the study of the network.

Suppose that the system is initiated at time t_0 , and let $t \geq t_0 + m + 2$. Then $r_1(t-1) = r_{m+2}(t-m-2)$ by (9). But then $a(t-m-1) = r_{m+2}(t-m-2) = r_1(t-1)$, by (10). By Lemma 5, $c(t-1)$ is equal to either $r_1(t-1)$ or $g(r_1(t-1))$, depending on whether $d(t-m-1) = (0, 1)$ or $(1, 0)$ respectively. Since $r_{N+1}(t) = c(t-1)$, we have

$$r_{N+1}(t) := \begin{cases} r_1(t-1), & \text{if } \text{mode}(t-1) = \text{rotate}, \\ g(r_1(t-1)), & \text{if } \text{mode}(t-1) = \text{switch}, \end{cases} \quad (11)$$

where the auxiliary Boolean variable $mode(t)$ is defined as $mode(t) := switch$ if $d(t-m) = (0, 1)$ and $mode(t) := rotate$ if $d(t-m) = (1, 0)$. The variable $mode$, similarly as r_1, \dots, r_{m+1} , is defined merely for the purposes of this proof, and it does not form part of the network itself.

Suppose now that $t_0 \leq -m - \mu - 2$. At time 0, assume that $r_{N+1} = \text{START}$, and $r_\eta \neq \text{START}$, for $1 \leq \eta \leq N$. Let $j \geq 0$ be such that $r_\eta = \text{ACTIVE}$ for $1 \leq \eta \leq j < N$, and $r_{j+1} \neq \text{ACTIVE}$. We show that

$$mode(\eta) = switch, \quad 0 \leq \eta \leq j; \quad mode(\eta) = rotate, \quad j+1 \leq \eta \leq N. \quad (12)$$

To see this, note that by (9) $r_{m+\mu+2}(\eta) \neq \text{START}$, for $-m - \mu - 1 \leq \eta \leq N - m - \mu - 2$. Since $g^{-1}(\text{START}) = \text{START}$, it also follows that $\text{START} = r_1(0) = c(-1) = a(-m-1) = r_{m+2}(-m-2)$, and $r_{m+\mu+2}(-m-\mu-2) = \text{START}$. Thus setting $t = -m - \mu - 1$, one has $p(t) = \text{START}$, $p(\eta) \neq \text{START}$ for $t < \eta \leq t + N$, $p(\eta) = \text{ACTIVE}$ for $t+1 \leq \eta \leq t+j$, and $p(t+j+1) \neq \text{ACTIVE}$. Applying Lemma 6 with $M := N$, we have that $q(\eta) = (1, 0)$, for $t+\mu \leq \eta \leq t+\mu+j$, and $q(\eta) = (0, 1)$, $t+\mu+j < \eta \leq t+\mu+N$. Equation (12) then follows directly from the definition of d and the mode variable. It is analogous to verify that (12) also holds in the case $j = N$, i.e. when $r_\eta(0) = \text{ACTIVE}$ for $1 \leq \eta \leq N$.

Note that using equation (12) we can fully calculate $r(N+1)$, namely $r_\eta(N+1) = g(r_\eta(0))$, for $1 \leq \eta \leq j+1$, and $r_\eta(N+1) = r_\eta(0)$ for $j+1 < \eta \leq N$; also, necessarily $r_{N+1}(N+1) = \text{START}$ regardless of $mode(N)$, since $g(\text{START}) = \text{START}$. The same process can be repeated starting at time $N+1, 2(N+1)$, etc., since necessarily $r_\eta = \text{START}$ can still only hold for $\eta = N+1$.

An appropriate initial condition to reach the above situation can be given as follows. Let $t_0 = -(N+1)$, and let $r_\eta(t_0) = \chi^{-1}(0, \dots, 0)$, for $1 \leq \eta \leq N$. Let $r_{N+1}(t_0) = \text{START}$. Finally, let $B(D)$ be initialized with $m(\mu)$ successive inputs of $a = \chi^{-1}(0, \dots, 0)$, $d = (0, 1)$ ($p = \chi^{-1}(0, \dots, 0)$). This way for $t = 0$ we guarantee that $r_{N+1} = \text{START}$, $r_\eta(t_0) = \chi^{-1}(0, \dots, 0)$ for $1 \leq \eta \leq N$, and importantly, $t_0 \leq -m - \mu - 2$.

Finally, under our standing hypotheses $t_0 \leq -m - \mu - 2$, $r_{N+1}(0) = \text{START}$, and $r_\eta(0) \neq \text{START}$ for $1 \leq \eta \leq N$. Define the following initial conditions for the system (5), (6): $s_\eta(0) := \chi^{-1}(r_\eta(0))$, $i = 1 \dots N$. After calculating j as before, $0 \leq j \leq N$, we have seen that $s_\eta(N) = \Gamma(s_\eta(0))$ for $1 \leq \eta \leq j+1$, and $s_\eta(N) = s_\eta(0)$ otherwise. From the discussion above, it follows that $\chi^{-1}(r_\eta(N+1)) = s_\eta(N)$ for $1 \leq \eta \leq N$. This equivalence between the two systems implies in particular that the states $r(t)$ are pairwise different for $t = 0, N+1, 2(N+1), \dots, (2^{lN} - 1)(N+1)$. The result follows. \square

2.1 Proof of Theorem 1

We can use Proposition 7 and Theorem 8 to prove the theorem stated in the introduction. Let $0 < c < 2$ be arbitrary. We prove first that there exist $L > 0$ even and integer $l > 0$ such that

$$\binom{L}{L/2} > 2^l > c^L. \quad (13)$$

The second inequality is equivalent to $L/l < \ln 2 / \ln c$; thus let $L = wl$, for some fixed $1 < w < \ln 2 / \ln c$ (for large enough l , L can then be rounded up to the nearest even number while satisfying this inequality). Using Stirling's formula, we have $\binom{L}{L/2} > v 2^L / \sqrt{2\pi L}$ for large enough L , where $0 < v < 1$ is arbitrary and fixed. The first inequality in (13) is satisfied if $v 2^L / \sqrt{2\pi L} > 2^l$. But after replacing $L = wl$ this is equivalent to $2^{(w-1)l} > v^{-1} \sqrt{2\pi wl}$. Clearly this inequality is satisfied for sufficiently large l , hence (13) follows.

The first inequality is now used to carry out the construction of system S , which by Theorems 7 and 8 is cooperative and bi-quadratic with strongly connected digraph, and has a periodic orbit of length greater than or equal to 2^{Nl} .

It remains to show that $2^{Nl} \geq c^n$ for large $N > 0$, where n is the dimension of the system. Let T be the total number of variables in the subnetworks D, B . Note that T depends only on L, l , and not on N . Then $n = (N + 1 - (m + 1))L + T = NL - mL + T$. Notice that $c^n \leq 2^{Nl}$ if and only if $(NL - mL + T) \ln c \leq Nl \ln 2$, which holds if and only if

$$L \ln c \leq l \ln 2 + \frac{mL - T}{N} \ln c.$$

But this equation is satisfied for large enough N , since $L \ln c < l \ln 2$ by (13). \square

References

- [1] Albert R. and Othmer, H.G. (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*, *J. Theor. Biol.* 223:1-18.
- [2] Angeli, D. and Sontag, E.D. (2003) Monotone control systems, *IEEE Trans. Autom. Control* 48:1684-1698.
- [3] Angeli, D., Ferrell, J.E., and Sontag, E.D. (2004) Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems, *Proc. Nat. Acad. Sci.* 101:1822-1827.
- [4] Arnone, M. I. and Davidson, E. H. (1997) The hardwiring of development: organization and function of genomic regulatory systems, *Development* 124:1851-1864.
- [5] Colon-Reyes, O., Laubenbacher, R., and Pareigis, B. (2004) Boolean monomial dynamical systems, *Ann. Comb.*, 8(4):425-439.
- [6] Dee, D. and Ghil, M. (1984) Boolean difference equations, I: formulation and dynamic behavior, *SIAM J. Appl. Math.* 44(1):111-126.

- [7] Garcia, L., Jarrah, A.S., and Laubenbacher, R. (2006) Sequential dynamical systems over words, *Appl. Math. Comput.* 174:500–510.
- [8] Gilbert, E.N. (1954) Lattice theoretic properties of frontal switching functions, *J. Math. Phys.* 33:57-67.
- [9] Greil, F., and Drossel, B. (2007) Kauffman networks with threshold functions, *Europ. Phys. J. B* 57:109–113.
- [10] Hernandez Toledo, R.A. (2005) Linear finite dynamical systems, *Communications in Algebra* 33:2977-2989.
- [11] Just, W. and Enciso, G. (2007) Analogues of the Smale and Hirsch theorems for cooperative Boolean and other discrete systems, *preprint*, arXiv:math/0711.0138v1.
- [12] Kauffman, S.A. (1969) Homeostasis and differentiation in random genetic control networks, *Nature* 224:177-178.
- [13] Kauffman, S.A. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets, *J. Theor. Biol.* 22(3):437-467.
- [14] Kauffman, S. A. (1993) *Origins of Order: Self-Organization and Selection in Evolution*. Oxford U Press, 1993.
- [15] Smith, H.L. (1995) *Monotone dynamical systems*, Math Surveys and Monographs, AMS, Providence, RI.
- [16] Sontag, E.D., Veliz-Cuba, A., Laubenbacher, R., and Jarrah, A.S. (2007) The effect of negative feedback loops on the dynamics of Boolean networks, *preprint*, arXiv:0707.3468v2 [q-bio.QM]
- [17] Sontag, E.D. (2007) Monotone and near-monotone biochemical networks, *Journal of Systems and Synthetic Biology*, to appear.
- [18] A.H.Y. Tong et al. (2004) Global mapping of the yeast interaction network, *Science* 303:808-813.
- [19] Tosić, P.T., and Agha, G. (2004) Characterizing configuration spaces of simple threshold cellular automata, Springer-Verlag LNCS series, vol. 3305, pp. 861 - 870.